Information theory of massively parallel probe storage channels.

Oliver Hambrey, Thomas Parnell and Oleg Zaboronski

Abstract—**Motivated by the concept of probe storage, we study the problem of information retrieval using a large array of N nano-mechanical probes, $N \sim 4000$. At the nanometer scale it is impossible to avoid errors in the positioning of the array, thus all signals retrieved by the probes of the array at a given sampling moment are affected by the same amount of random position jitter. Therefore a massively parallel probe storage device is an example of a noisy communication channel with long range correlations between channel outputs due to the global positioning errors.**

We find that these correlations have a profound effect on the channel's properties. For example, it turns out that the channel's information capacity does approach 1 bit per probe in the limit of high signal-to-noise ratio, but the rate of the approach is only polynomial in the channel noise strength. Moreover, *any* error correction code with block size $N >> 1$ such that codewords correspond to the instantaneous outputs of the all probes in the array exhibits an error floor independently of the code rate. We illustrate this phenomenon explicitly using Reed-Solomon codes the performance of which is easy to simulate numerically.

We also discuss capacity-achieving error correction codes for the global jitter channel and their complexity.

Index Terms—**Probe storage, maximum likelihood detection, Shannon capacity, Gallager random coding bound, error exponent, Fano inequality**

I. INTRODUCTION

THE invention of atomic force microscopy in 1986 [1] opened the possibility of storing information at nanometer scales resulting in proposals for achieving aerial information densities of tens or even hundreds of Terabits per square inch. The basic idea is that information can be stored by altering certain features of the storage medium at the scale of nanometers. These changes can be then sensed and information retrieved by nanoscale probes similar to those used in atomic force microscopy. For example, binary information can be stored in crystalline dots created in amorphous media or amorphous dots in crystalline media and retrieved using electric probes (phase change storage, [2]). Alternatively, the information can be stored using indentations or, more recently, variable length grooves made in polymer media and retrieved using thermoelectric probes (thermo-mechanical probe storage, [3]).

O. Hambrey is with the Complexity Doctoral Training Centre, University of Warwick, Coventry, CV4 7AL, UK and Siglead Europe Ltd., International Digital Laboratory, University of Warwick, Coventry, CV4 7AL, UK e-mail: (oliver.hambrey@siglead.com)

T. Parnell is with Siglead Europe Ltd., UK e-mail: (tom.parnell@siglead.com)

O. Zaboronski is with the Warwick Mathematics Institute, University of Warwick and Siglead Europe Ltd. e-mail: (O.V.Zaboronski@warwick.ac.uk)
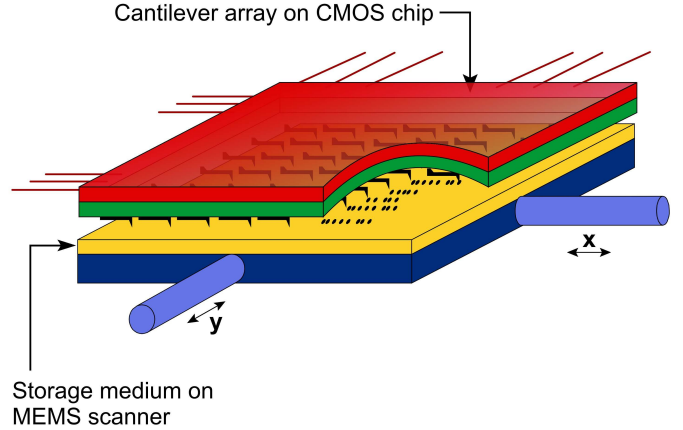

Fig. 1. The layout of the 'Millipede' - a thermomechanical probe storage device. Courtesy of IBM - Zurich, [7]

The whole concept of storing and retrieving information using nano-scale probes became known as probe storage, see [4], Chapter 4, for a comprehensive review of the current state of the field. Perhaps the most widely known concept of the probe storage device is the IBM's 'Millipede' [8], [5] - an array of thermo-electric probes with sharp tips used to create and sense indentations in the polymer media. The layout of the Millipede is shown in Fig. 1, the basic principle of thermo-mechanical reading and writing is explained in Fig. 2. It has been demonstrated that data can be retrieved with low bit error rate per probe (about 10^{-4}) at densities of up to 2 Terabit per square inch [6].

The particular feature of the Millipede shared by all existing concepts of probe storage is the presence of a large ($\sim 2^{12}$) number of probes reading and writing the information in parallel. This feature makes probe storage very different from more traditional storage devices such as magnetic or optical disks or flash memory. Each probe reads and writes information in its own field. The array of probes moves as a whole to allow each probe to explore its field. For aerial densities of information of several terabit per square inch, the array has to be moved by a distance of the order of 10 nanometers from one set of the sampling points to the next and repositioned with a sub-nanometer precision for writing or reading. Inevitably, positioning errors affect every single probe in the array. We refer to the combined effect of errors in the positioning of the array at both the reading and writing stages as global positioning jitter or just global jitter. The aim of this paper is to investigate the performance of error correcting codes operating on the output of the probe array subject to highly correlated
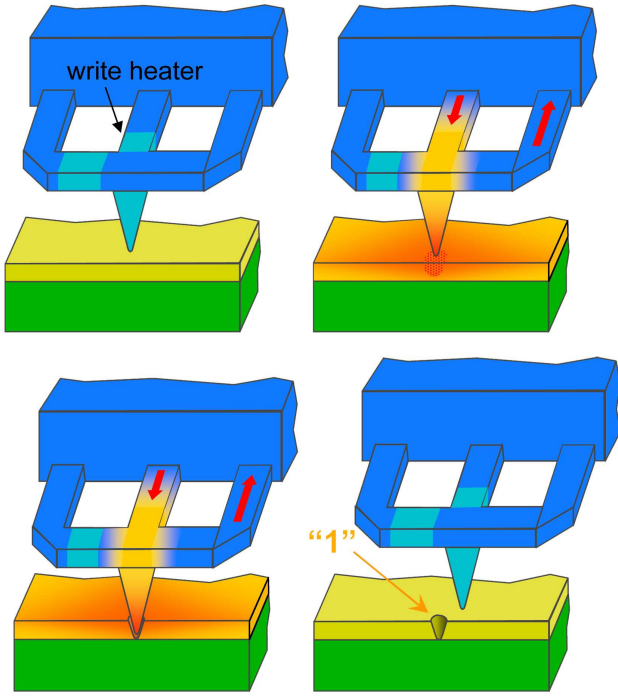
Fig. 2. The principle of thermomechanical reading and writing. Top left picture, writing bit 'zero': cold probe pressed against the polymer surface leaves no mark in the media. Top right picture, writing bit 'one': the probe heated above the polymer's melting temperature and pressed against the polymer surface leaves an indentation in the media shown in the bottom left picture. Bottom right pictures, reading bits 'one' and 'zero': the probe inserted in the indentation is cooler than the probe pressed against the flat surface. These temperature variations can be captured using thermo-resistive sensors. Courtesy of IBM Zurich, [8]

disturbances due to global jitter. We will investigate both the performance of special codes (Reed-Solomon) and address the general question of existence of good error correction codes for the global jitter channel by calculating channel capacity and studying Gallager's random coding bound.

Despite the fairly abstract tools used in the paper, the purpose of our investigation is to answer a very practical question: given a 'good' communication channel (e. g. the single probe's channel with bit error rate 10^{-4}), how good is a communication system (e.g. a Millipede) consisting of thousands of good channels subjected to correlated error events?

Our system-level analysis allows one to understand advantages and limitations of complicated communications devices without actually building one. In particular, it turns out that information theoretical performance limits derived in this paper have crucial system-level implications for the design of error correction codes for probe storage.

The rest of the paper is organised as follows. In Section II we introduce a simple model for the probe storage channel, the global jitter channel, which accounts for the effects of the probe array's position jitter and calculate the probability distribution function of signal amplitude for Gaussian isolated pulse response and Gaussian statistics of jitter. In Section III we introduce a low complexity signal detection scheme for the global jitter channel and prove its optimality in the limit of

large array sizes. In Section IV we calculate block error rate for non-interleaved Reed-Solomon codes applied to the probe storage channel and analyze their error floor behaviour. In Section V we calculate Shannon's capacity of the global jitter channel and show that it approaches 1 bit according to a power law in the limit of large signal-to-noise ratio. In Section VI we calculate the average block error rate for non-interleaved codes sampled from Gallager's ensemble (the random coding bound or RCB) and show that it exhibits an error floor behaviour as a function of SNR. In Section VII we show that there exists no non-interleaved codes with a positive rate which can be used for error-free retrieval of information using large arrays of probes in the presence of global jitter. In Section VIII we discuss minimal requirements for capacity-achieving codes for global jitter channels. We conclude our work with Section IX which contains the summary of the results of our investigation. For completeness, we present the derivations of more technical results obtained in the paper in the Appendix.

II. CHANNEL MODEL

We consider the system consisting of an array of N probes reading/writing in parallel. We assume that channel coding has been used (an RLL code for example) and that the symbol pitch is large enough so that inter-symbol interference can be ignored. The sampled readback signal at the k-th probe at the t-th moment in time is modelled as:

$$r_t^{(k)} = p(J_t)a_t^{(k)} + \sigma n_t^{(k)}, \ k = 1, 2, \ldots, N \quad (1)$$

where $a_t^{(k)} \in \{0, 1\}$ is the bit written to the medium by the k-th probe at time t, J_t is the global positioning error (jitter) at time t, $p(J)$ is the channel impulse response and $\{\sigma n_k^{(t)}\}$ is a sequence of random variables modelling the combined effect of electronics and media noise.

Experiments carried out at IBM Zurich [9] have confirmed that for thermo-mechanical storage media the combined electronics/media noise is well modelled by Gaussian random variables. Thus we assume that $\{n_t^{(k)}\}$ are independent identically distributed Gaussian variables with mean zero and unit variance. Parameter σ is the standard deviation of the resulting additive white Gaussian noise (AWGN).

It has also been demonstrated experimentally that when using a Millipede-like positioning control loop the random jitter J_t is also well modeled by a mean-zero Gaussian random variable [9]. Note that this observation applies to any probe storage device, for instance a system based on phase change media, as long as a similar positioning system is used. Let σ_J be the standard deviation of Gaussian jitter.

Most of the qualitative results reported below do not depend on the detailed assumptions about the statistics of jitter or the precise shape of the impulse response. For quantitative analysis and numerical simulations of the global jitter channel we will use the Gaussian impulse response:

$$p(J) = e^{-\frac{J^2}{W^2}} \quad (2)$$

where W is a parameter related to pulse width. In what follows it will be more convenient to work directly with the random variable $p(J) \in (0, 1)$ which measures signal amplitude

degradation due to position jitter J. A calculation leads to the following answer for the probability density function of p:

$$\rho_P(p) \equiv \mathbf{E}_J \delta(p - p(J)) = \frac{W}{\sqrt{2\pi\sigma_J^2 \log\left(\frac{1}{p}\right)}} p^{\frac{W^2}{2\sigma_J^2} - 1}. \quad (3)$$

Finally we define the signal-to-noise ratio (SNR) of the global jitter channel measured in decibels as follows:

$$SNR = 10 \log_{10}\left(\frac{1}{\sigma^2}\right) \quad (4)$$

III. OPTIMAL CHANNEL DETECTOR

The global jitter channel is characterised by strong equal time correlations between outputs of all N-probes in the array. In the limit $N \to \infty$ it is possible to exploit these correlations and derive an asymptotically optimal low-complexity detection scheme.

To motivate the rigorous argument given below, let us ask the following question: what is the optimal channel detector conditional on the knowledge of the value of jitter at time t? Conditional on the known amplitude value $p_t \equiv p(J_t) = p$, channel outputs $r_t^{(k)}$'s are independent. Consequently, the optimal maximum a posteriori (MAP) detector [11] is simply the collection of N independent optimal threshold detectors for AWGN channel:

$$\hat{a}_t^{(k)} = \begin{cases} 1 & \text{if } r_t^{(k)} > \frac{p}{2} \\ 0 & \text{if } r_t^{(k)} < \frac{p}{2} \end{cases} \quad (5)$$

The problem with the above 'Genie-assisted' detector is that the value of the amplitude p is a priori unknown. However the value of p can be reliably estimated from the string $r_t^{(1)}, r_t^{(2)}, \ldots, r_t^{(N)}$ if N is sufficiently large. The key observation concerns the sample average of all N received signals at time t:

$$\begin{aligned} \overline{R}_t &= \frac{1}{N} \sum_{k=1}^{N} r_t^{(k)} \\ &= p(J_t) \sum_{k=1}^{N} \frac{a_t^{(k)}}{N} + \sigma \sum_{k=1}^{N} \frac{n_t^{(k)}}{N} \end{aligned} \quad (6)$$

Assume that channel inputs $a_t^{(k)}$ are independent and uniformly distributed. Then by the strong law of large numbers [10], both sample averages in (6) converge almost surely to their expected values as $N \to \infty$:

$$\sum_{k=1}^{N} \frac{a_t^{(k)}}{N} \xrightarrow{a.s.} \frac{1}{2} \qquad \sum_{k=1}^{N} \frac{n_t^{(k)}}{N} \xrightarrow{a.s.} 0 \quad (7)$$

Substituting into (6) we find in the large N limit it is actually possible to compute the previously unknown distortion $p_t = p(J_t)$ as follows:

$$p_t = \lim_{N \to \infty} \left(2\overline{R}_t\right) \quad (8)$$

Feeding this estimate of p to N independent threshold detectors (5) we obtain estimates $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N$ of data bits which should become optimal in the limit $N \to \infty$.

The heuristic argument given above leads to the following asymptotically optimal detection algorithm for the global jitter channel:

1) **Input:** Channel output $r_t^{(1)}, r_t^{(2)}, \ldots r_t^{(N)}$ at time t.
2) Estimate the threshold:

$$T_N = \frac{1}{N} \sum_{k=1}^{N} r_t^{(k)} \quad (9)$$

3) Perform bit-by-bit detection:

$$\hat{a}_t^{(k)} = \begin{cases} 1 & \text{if } r_t^{(k)} > T_N \\ 0 & \text{if } r_t^{(k)} < T_N \end{cases} \quad (10)$$

4) **Output:** The estimate of channel input $\hat{a}_t^{(1)} \hat{a}_t^{(2)} \ldots \hat{a}_t^{(N)}$.

We will now prove that detector (10) is indeed optimal in the limit $N \to \infty$. Firstly, note that detector (5) is an optimal MAP detector which infers the most probable channel input conditional on $\mathbf{r_t}, \mathbf{p_t}$. Therefore its bit error rate is smaller or equal than bit error of any detector which infers channel input conditional on channel input $\mathbf{r_t}$ only. So we can establish the asymptotic optimality of (10) by proving that its bit error rate approaches the bit error rate of (5) in the limit $N \to \infty$:

$$\lim_{N \to \infty} |\Pr(\hat{a}_t^{(k)} \neq a_t^{(k)}) - \Pr(\hat{a}_t^{(k)} \neq a_t^{(k)})| = 0. \quad (11)$$

A long but straightforward calculation presented in Appendix A shows that

$$|\Pr(\hat{a}_t^{(k)} \neq a_t^{(k)}) - \Pr(\hat{a}_t^{(k)} \neq a_t^{(k)})| \leq \frac{3\left(1 + \frac{\mathbb{E}(p^2)}{4\sigma^2}\right)^{1/3}}{(2\pi N)^{1/3}}. \quad (12)$$

Taking the large-N limit of both sides of the above inequality we arrive at (11). Thus the asymptotic optimality of (10) is established. We will refer to detector (10) as the LLN-detector and detector (5) as $Genie - detector$.

The performance of the LLN-detector for large but finite N is compared with the performance of Genie-detector in Fig. 3 for Gaussian pulse (2) and various values of jitter strength σ_J. The size of the probe array is $N = 1000$, the pulse width is $W = 0.5$ and the simulation has been run for 10^5 consecutive readings of the entire probe array. As a reference the performance of the channel with no jitter which corresponds to the classical binary-input additive white Gaussian noise (AWGN) channel has been included. It is observed that as jitter strength is increased the BER performance degrades significantly but for all cases the performance of the LLN-detector is virtually indistinguishable from Genie-assisted detector. It is interesting to note that the gap between BER of Genie-detector and the LLN-detector increases with SNR, see Fig. 3. This observation is consistent with the bound (12): The right hand side of the bound grows if the block size N is kept fixed but the additive noise strength σ is reduced. The intuition behind the observed divergence is very simple: for smaller noise one needs a more precise threshold estimate to stay near the performance of the Genie-detector. The precision of the threshold estimate depends on the rate of convergence of the sample sum and scales as $1/\sqrt{N}$ in accordance with Central Limit Theorem (see [10] for a review).
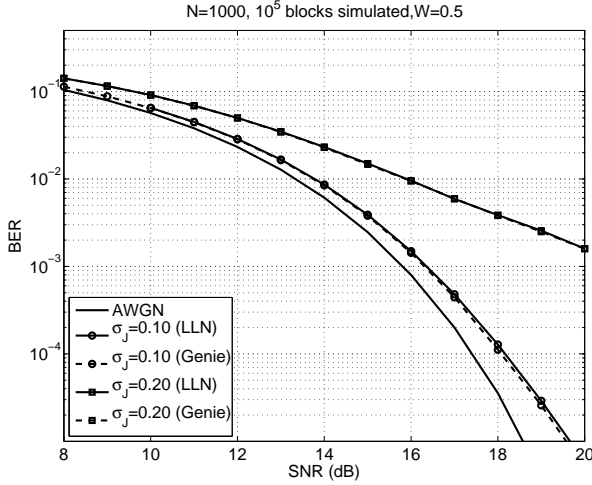
Fig. 3. BER performance comparison of LLN detector against the ideal detector with perfect knowledge of global jitter

Fig. 4. Comparison of BER performance of the LLN-detectors for the global jitter channel and the set of N-independent threshold detectors

It is interesting to compare the LLN detector to the simplest detection scheme proposed for Millipede, see [12]. The Millipede detector consists of N independent MAP detectors (one detector per probe). The k-th detector estimates the data bit $a_t^{(k)}$ at time t conditional on the output $r_t^{(k)}$ only. Unlike the LLN-detector, there is no sharing of information between individual detectors.

The maximum-likelihood detector of the single probe output modeled by (1) is a simple threshold detector: the most likely bit given the output of the k-th probe, $\hat{a}_t^{(k)}$ is given by:

$$\hat{a}_t^{(k)} = \begin{cases} 1 & \text{if } r_t^{(k)} > r_0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where r_0 is the optimal threshold - a number between 0 and 1 which can be either measured experimentally or computed theoretically for a given the channel model by solving the maximum likelihood equation

$$\Pr(r_t^{(k)} = r_0 \mid a_t^{(k)} = 1) = \Pr(r_t^{(k)} = r_0 \mid a_t^{(k)} = 0),$$

see [13] for more details.

In Figure 4 the BER of the Millipede detection scheme is compared with that of the LLN-detector for two values of jitter strength σ_J. The result is quite striking: the optimal detector for the global jitter channel outperforms the set of N independent threshold detector by over a decibel for BER$\sim 10^{-4}$.

Finally, let us analyse the complexity of the LLN-detector. The complexity of adding N fixed precision numbers in (9) scales as $N log(N)$, the complexity of the detection step (10) is $O(N)$ so the overall detection complexity is $O(N log(N))$, the detection complexity per detected bit is $O(log(N))$.

Having constructed the optimal detection scheme we can investigate the performance of error correction codes for the global jitter channel starting with Reed-Solomon codes which featured in the original Millipede proposal.
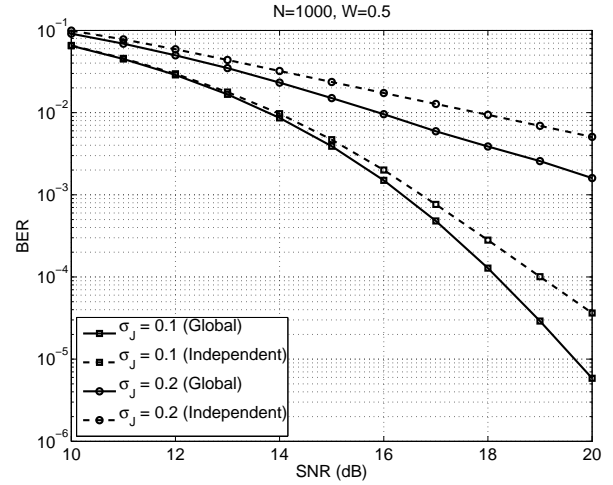
IV. THE PERFORMANCE OF REED-SOLOMON CODES FOR THE GLOBAL JITTER CHANNEL.

The typical probe storage array size N is $64 \times 64 = 4096$, which is close to the sector size of the previous generation of hard disk drives.[1] The simplest error correction coding (ECC) scheme for probe storage follows the example of hard drives: for the latter data is encoded sector-by-sector, for the former ECC is applied independently to K-bit strings of data $\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3 \ldots$ to be recorded on the media by the probe array at times $t_1, t_2, t_3, \ldots$. We refer to the described application of error correction coding as *non-interleaved* meaning that channel outputs corresponding to different moments of sampling time cannot belong to the same code block.

The non-interleaved ECC block size is equal to the number of probes N, the ECC rate is $R = K/N$. During the reading stage, the single-time output $\hat{a}_t^{(1)}, \hat{a}_t^{(2)}, \ldots, \hat{a}_t^{(N)}$ of the channel detector is fed into the ECC decoder resulting in an estimate $\hat{\mathbf{I}}_t$ of K recorded bits at each sampling time t.

We start our study of error correction for global jitter channel with classical Reed-Solomon (RS) codes, see [14] for review. These are the (N_s, K_s) symbol block codes over the Galois field $GF(2^n)$. Here $N_s = N/n$ is the number of RS symbols per block, $K_s = K/n$ is the number of information symbols. Symbols are represented by n-bit binary strings. The maximal block size of the code is $N_{max} = 2^n - 1$ symbols. An RS code with any block size $N_s \leq N_{max}$ can be constructed by treating missing $(N_{max} - N_s)$ symbols as zeros (this operation is called shortening). A Reed-Solomon code with rate $R = K_s/N_s$ can correct up to

$$N_s \frac{1-R}{2}$$

incorrectly detected symbols, which makes it a maximal distance separable (MDS) code.

---

[1]January 2011 has been designated as the date of the final transition from 512B sector size in HDD's to the 4KB sector size.

The fact that the error event of a RS code depends on the number of incorrectly detected symbols rather than bits, makes it a very good code for channels dominated by relatively short bursts of noise. It is therefore not very well suited for the global jitter channel. However we will see below that conclusions drawn from analyzing the performance of RS codes in the presence of global jitter can be applied to *any* non-interleaved block ECC.

We start with presenting results of numerical simulations. Fig. 5 illustrates the performance of rate-0.8 RS codes with symbol size $n = 10$ bits. Three codes are considered: with block size $N_s = 1023$ (the maximal block size) and two shortened RS codes - with $N_s = 511$ and $N_s = 255$. These block sizes correspond to hypothetical probe arrays with 2550, 5110 and 10230 probes. To model global jitter we use the channel (1) with jitter strength $\sigma_j = 0.2$, and pulse width $W = 0.5$. For each of the codes we measure the probability of sector error or sector error rate (SER) as a function of signal-to-noise ratio (SNR). SER is measured as the number of times the number of symbol errors in the received string exceeded RS threshold $N_s\frac{1-R}{2}$. The total number of strings is 10^6 to ensure that the total error count used to estimate SER is at least 10^2 for the highest SNR point.

The first striking feature of the curves presented in Fig. 5 is that there no discernible performance loss using the shortened codes. This is in stark contrast from the known behaviour of RS codes for the AWGN channel where

$$\frac{logSER_{RS(N_1,R)}}{logSER_{RS(N_2,R)}} \xrightarrow{N \to \infty} \frac{N_1}{N_2}.$$

Therefore, the numerical evidence suggests that the probability of RS codes applied to the global jitter channel does *not* depend on the code's block size.

The second unusual feature of SER vs. SNR curves shown in Fig. 5 is the exponential law of the decay of SER with SNR - all the curves look like straight lines on the semilogarithmic plot. The exponential rather than the 'waterfall' shape of SER curves is often referred to as an 'error floor'. Its appearance is normally attributed to the use of suboptimal decoding algorithms such as belief propagation, rather than channel properties. Here we are driven to a conclusion that non-interleaved RS codes decoded with an optimal maximum likelihood hard input algorithm exhibit an error floor, the position of which is independent of the block size.

As it turns out, the breakdown of Reed-Solomon codes for channels with global jitter can be understood analytically using the machinery of large deviations, [16].

To achieve this, we need to introduce some notations. Let

$$\tau = (1 - R)/2 \tag{14}$$

be the maximal fraction of correctable symbols for the code at hand. Let the random variable ξ_k take values $\{0, 1\}$ with probabilities $\{e_0, e_1\}$ respectively. Here e_1 is the probability that the k-th symbol is detected incorrectly. Due to statistical homogeneity of the channel model (1) the probability of symbol error does not depend on the symbol index k. The event $\xi_k = 1$ corresponds to the k-th symbol being detected erroneously and $\xi_k = 0$ the k-th symbol being detected

correctly. Conditionally on the event $p_t = p$ the probability of RS sector error is simply given by a multinomial formula [15]. Unfortunately, the multinomial formula is not very useful for quantitative analysis in the region of low SER. Instead, we are going to use simple asymptotic expressions for RS sector error rate based on Cramer's theory [16]. The event of sector error corresponds to the fraction of incorrectly detected symbols exceeding RS threshold (14). Hence, the probability of a block being decoded incorrectly can be written:

$$\Pr(SE \mid p_t = p) = \Pr\left(\sum_{k=1}^{N_s} \xi_k > \tau N_s \mid p_t = p\right) \tag{15}$$

An application of Chernoff's bound [10] results in the following upper bound on $\Pr(SE)$ for any $\lambda > 0$:

$$\Pr(SE \mid p_t = p) \le e^{-\lambda \tau N_s} \mathbb{E}\left[e^{\lambda\left(\sum_{k=1}^{N_s} \xi_k\right)} \mid p_t = p\right], \tag{16}$$

where $\mathbb{E}[\bullet \mid p_t = p]$ stands for p_t-conditional expectation value. Recall that conditional on the value of signal amplitude p_t the global jitter channel is memoryless. Therefore symbol error events ξ_k are conditionally independent and identically distributed and the bound in (16) can be re-written as:

$$\Pr(SE \mid p_t = p) \le e^{-\lambda \tau N_s + N_s \ln \mathbb{E}[e^{\lambda \xi} \mid p_t = p]} \tag{17}$$

The bound in (17) holds for any $\lambda > 0$ and thus we can choose the tightest bound possible by minimising over all λ's:

$$\frac{1}{N_s} \ln \Pr(SE \mid p_t = p) \le I(\tau, p) \tag{18}$$

Function $I(\tau, p)$ is known in the theory of large deviations as the rate function and is given by:

$$I(\tau, p) = \inf_{\lambda > 0}\left(-\lambda\tau + \ln \mathbb{E}\left[e^{\lambda\xi} \mid p_t = p\right]\right) \tag{19}$$

The question remains: is the bound in (18) tight enough to be useful? The answer is provided by an application of Cramer's Theorem [16] which states that provided $\mathbb{E}[\xi \mid p_t = p] < \tau$ (which is certainly true in the limit of low sector error-rates) the bound (18) is tight in the limit of large block size:

$$\lim_{N_s \to \infty} \frac{1}{N_s} \ln \Pr(SE \mid p_t = p) = I(\tau, p) \tag{20}$$

An explicit expression for the rate function (19) can be found by solving the critical point equation:

$$\frac{\partial}{\partial \lambda}\left(-\lambda\tau + \ln \mathbb{E}\left[e^{\lambda\xi} \mid p_t = p\right]\right) = 0 \tag{21}$$

The function differentiated in (21) is convex with the unique point of global minimum given by:

$$\lambda_c = \ln\left(\frac{e_0(p)}{e_1(p)}\frac{\tau}{(1 - \tau)}\right), \tag{22}$$

where $e_1(p) = \mathbb{E}(\xi \mid p_t = p)$ is the conditional symbol error rate, $e_0(p) = 1 - e_1(p)$. If the condition of Cramer's theorem $\mathbb{E}[\xi \mid p_t = p] = e_1(p) < \tau$ is satisfied then $\lambda_c > 0$ and substituting back into (19) we find the rate function can be expressed as a Kullback-Leibler divergence [17]:

$$I(\tau, p) = -D_{KL}\left((1 - \tau, \tau) \,\|\, (1 - e_1(p), e_1(p))\right) \tag{23}$$

Recall that for any two stochastic vectors $\mathbf{P}$ and $\mathbf{Q}$,

$$D_{KL}(\mathbf{P}||\mathbf{Q}) = \sum_k P_k \log\left(\frac{P_k}{Q_k}\right) \qquad (24)$$

On the other hand if $e_1(p) \geq \tau$ then $\lambda_c \leq 0$ and due to convexity the minimum of (19) is achieved at $\lambda = 0$ resulting in the trivial bound $I(\tau, p) = 0$.

Let us summarise our findings so far:

$$\Pr(SE \mid p_t = p) \lesssim e^{-N_s I(\tau, p)}, \qquad (25)$$

where

$$I(\tau, p) = \begin{cases} D_{KL}\left((1-\tau, \tau) \,||\, (e_0(p), e_1(p))\right) & \text{if } e_1(p) < \tau \\ 0 & \text{if } e_1(p) > \tau \end{cases} \qquad (26)$$

Now we can derive an upper-bound for the performance of non-interleaved RS codes in the global jitter channel using the upper bound (25, 26) on the conditional SER:

$$\Pr(SE) = \int_0^1 d\mu_P(p) \, \Pr(SE \mid p_t = p), \quad (27)$$

where $d\mu_P(p) = \rho_P(p)dp$ is the probability measure of jitter-dependent signal amplitude with density (3). The range of integration can now be split around the critical value of signal degradation p_c which is the unique solution of the equation

$$e_1(p_c) = \tau.$$

Then using the upper bound (25) for the conditional probability of RS error we find:

$$\Pr(SE) \leq \int_0^{p_c} d\mu_P(p) + \int_{p_c}^1 d\mu_P(p) \, e^{-N_s F_\tau(p)} \qquad (28)$$

Where $F_\tau(p)$ is expressed in terms of Kullback-Leibler divergence:

$$F_\tau(p) = D_{KL}\left((1-\tau, \tau) \,||\, (e_0(p), e_1(p))\right)$$

Note that $F_\tau(p)$ has a unique non-degenerate minimum at p_c where it takes the value $F(p_c) = 0$. The proof of this fact follows easily from Gibbs inequality [17]. Therefore, we can apply the Laplace formula [18] to the second integral in equation (28) to derive the large-N_s asymptotic of the probability of sector error:

$$\int_{p_c}^1 d\mu_P(p) \, e^{-N_s \, F(p)} \sim \rho_P(p_c)\sqrt{\frac{\pi}{N_s F''(p_c)}} \qquad (29)$$

Note that the resulting expression (29) tends to zero as $\sqrt{\frac{1}{N_s}}$ in the limit $N_s \to \infty$. Therefore what we have discovered is that for the global jitter channel in the limit $N_s \to \infty$ the probability of sector error is upper-bounded as follows:

$$\Pr(SE) \lesssim \Pr(p_t \leq p_c) + \rho_P(p_c)\sqrt{\frac{\pi}{N_s F''(p_c)}} \qquad (30)$$

We therefore confirmed theoretically that for a fixed level of noise and in the limit of large sector size, the probability of sector error for information encoded with a Reed-Solomon code and transmitted over the global jitter channel does not depend on the code's block size. This conclusion is in perfect
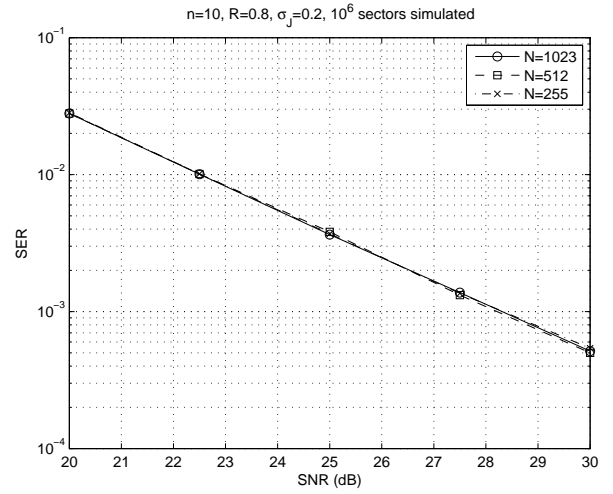

Fig. 5. RS SER curves do not depend on the block size and exhibit an error floor behaviour.

agreement with the results of numerical simulations shown in Fig. 5.

Using (30) we can also explain the shape of the SER vs SNR curves thus confirming the appearance of the error floor analytically. Moreover, we will be able to determine the position of the error floor as a function of code rate $R = 1-2\tau$.

Up to this point our considerations did not depend on the specific shape of the impulse response or on jitter statistics. From now on we will assume the Gaussian impulse response given by equation (2) and Gaussian position jitter. Then it follows from (30) that

$$\Pr(SE) \lesssim \Pr(p_t \leq p_c) + O(N_s^{-1/2})$$
$$= \frac{2}{\sqrt{2\pi\sigma_J^2}} \int_{J_c}^\infty dJ \, e^{-\frac{J^2}{2\sigma_J^2}} + O(N_s^{-1/2}) \qquad (31)$$

where $J_c = W\sqrt{\ln(p_c^{-1})}$ is the critical value of jitter that causes signal degradation p_c. Bounding the integral in equation (31) with elementary functions we arrive at the following:

$$\Pr(SE) \leq p_c^{\left(\frac{W^2}{2\sigma_J^2}\right)} \qquad (32)$$

If we further assume that a high-rate Reed-Solomon is used so that $\tau \ll 1$ then it follows that the symbol error rate conditional on p_c, $e_1(p_c) = \tau$ is also much less that 1. In this limit we can approximate symbol error rate as follows:

$$e_1(p_c) = 1 - (1 - f(p_c))^n \approx n f(p_c) \qquad (33)$$

where $f(p_c)$ is the bit error rate conditional on p_c which admits the following upper bound:

$$f(p_c) \leq \frac{1}{2} e^{-\frac{p_c^2}{8\sigma^2}} \qquad (34)$$

Note that this upper bound is tight in the limit of large signal to noise ratios - the region relevant for studying the error floor. Therefore we can estimate the critical point p_c by $\hat{p}_c \geq p_c$ as follows:

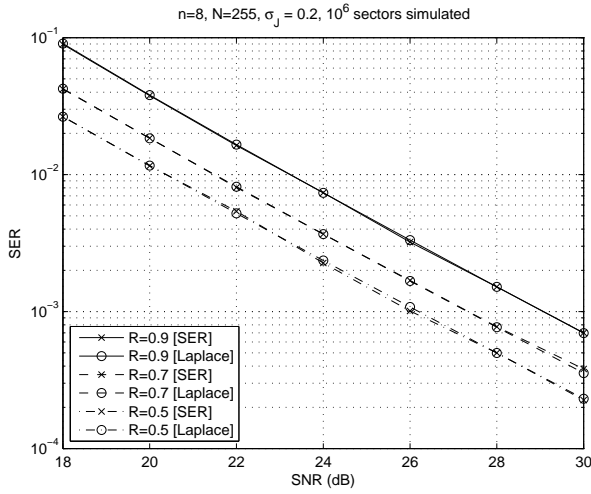$$\hat{p}_c = \sigma\sqrt{8\ln\left(\frac{n}{2\tau}\right)} \qquad (35)$$

Fig. 6. The rate of exponential decay of Reed-Solomon SER with SNR does not depend on the code rate.

Substituting (35) into equation (32) we arrive as the following bound on sector error rate for the global jitter channel with the exponential impulse response for high-rate Reed-Solomon codes:

$$\Pr(SE) \ \lesssim \ \left(8 \ln\left[\frac{n}{2\tau}\right]\right)^{\left(\frac{W^2}{4\sigma_J^2}\right)} \sigma^{\left(\frac{W^2}{2\sigma_J^2}\right)} + O(N^{-1/2}).$$
(36)

This is a disastrous result: the probability of sector error decays as power law in σ with an exponent that does not depend on code rate or block size! The algebraic dependence of SER on the noise strength implies the exponential dependence of the probability of sector error on signal-to-noise ratio which explains straight lines on the semi-logarithmic SER-SNR plot in Fig. 5.

Moreover, we observe that the position of the error floor (determined by the pre-factor in (36)) depends very weakly (logarithmically) on the code rate $R = 1 - 2\tau$.

In Figure 6 we compare numerical simulations of RS codes of various rates with expression (31) which is valid beyond the high rate approximation used to derive (36). We find a good agreement with our theoretical prediction: the rate of the exponential decay does not depend on either the block size or the code rate, the position of the error floor changes slowly with the code rate.

We conclude that non-interleaved Reed-Solomon codes applied to channel (1) exhibit an error floor: the probability of sector error decays exponentially as a function of SNR. Moreover, the rate of the exponential decay does not depend on either the code rate or the block size. The position of the error floor varies as a logarithm of the code rate. Non-interleaved Reed-Solomon codes are thus not suitable for use in a probe storage system that suffers from global positioning errors.

The above discussion suggests that the performance of Reed-Solomon codes can be improved with interleaving: by spreading the codewords over multiple time samples, the effect of occasional strong jitter leading to $p < p_c(\tau)$ can be

mitigated. We will discuss the complexity of this solution in Section VIII.

But first, we will address the following foundational question: do good codes for a global jitter channel exist in principle?

V. SHANNON CAPACITY OF THE GLOBAL JITTER CHANNEL.

Shannon's capacity C is one of the most important measures of quality for any communication channel. According to Claude Shannon's 1948 Channel Coding Theorem [19], there exists no error correction code of rate R which would achieve an arbitrarily small probability of error for a channel with capacity smaller than code rate, $C < R$. The 'positive' part of Shannon's theorem states that for any $R < C$ and for any $\epsilon > 0$ then there exists a block code C with block size $M(\epsilon)$ and rate less than or equal to R and a decoding algorithm such that the maximal probability of block error is less than ϵ. Thus by computing Shannon's capacity for the global jitter channel we will establish an upper bound on the rate of good error correction codes for this channel.

The definition of capacity rests on the notion of mutual information, see [17] for a review of fundamental notions of information theory. The mutual information between a discrete channel input $\mathbf{a}$ taking values in Ω_a and a continuous output ensemble $\mathbf{r}$ taking values in Ω_r is defined as follows:

$$I(R; A) = \sum_{\mathbf{a} \in \Omega_a} \Pr_A(\mathbf{a}) \int_{\Omega_r} d \Pr_{R|A}(\mathbf{r}|\mathbf{a}) \log_2 \frac{d\Pr_{R|A}(\mathbf{r}|\mathbf{a})}{d\Pr_R(\mathbf{r})}, \quad (37)$$

where

$$\frac{d\Pr_{R|A}(\mathbf{r}|\mathbf{a})}{d\Pr_R(\mathbf{r})}$$

is Radon-Nykodim derivative of the conditional probability measure $\Pr_{R|A}$ with respect to the marginal probability measure $\Pr_R$. If the probability densities $\rho_{R|A}(\mathbf{r} \mid \mathbf{a})$ and $\rho_R(\mathbf{r})$ of the probability measures $\Pr_{R|A}$ and $\Pr_R$ exist, the Radon-Nykodim derivative is simply the ratio of densities:

$$\frac{d\Pr_{R|A}(\mathbf{r}|\mathbf{a})}{d\Pr_R(\mathbf{r})} = \frac{\rho_{R|A}(\mathbf{r}|\mathbf{a})}{\rho_R(\mathbf{r})}.$$

Recall that for the global jitter channel, $\mathbf{a}$ is an N-bit data string and $\mathbf{r}$ is a string of N real valued signals generated by the probe array at a given sample time.

Shannon capacity is defined as the maximal mutual information over all input probability distributions per bit of input:

$$C = \max_{P_A} \frac{1}{N} I(R; A) \quad (38)$$

As it is easy to see from the definition, $0 \le C \le 1$ (bit). In particular, $C = 1$ for noiseless channels. No information can be communicated over the channel with $C = 0$ in finite time.

For a complicated channel where the maximisation over $\Pr_A$ is difficult to perform it is common to study a weaker form of capacity:

$$C_{i.u.d.} = \frac{1}{N} I(A; R)\Big|_{\Pr_A = \Pr_{i.u.d.}} \quad (39)$$

where $\Pr_{i.u.d.}$ is the probability distribution for which channel inputs are chosen independently and with equal probability. It is clear that

$$C_{i.u.d.} \leq C.$$

Therefore if we find that $C_{i.u.d.}$ is close to 1, then C is also close to one and we can be certain that there is a high rate ECC scheme which will achieve low probability of error for the global jitter channel. From this point onwards we will be only concerned with $C_{i.u.d.}$ and assume that $\Pr_A(\mathbf{a}) = \frac{1}{2^N}$, i. e. that all 2^N inputs sequences $\mathbf{a} = (a_1, \ldots, a_N)$ are sampled independently according to a uniform distribution.

The calculation of channel capacity is a notoriously difficult problem. Analytically it can be evaluated for the very simplest channels only, such as binary symmetric, binary erasure or AWGN channels [17], Chapter II. There exist efficient numerical algorithms for calculating capacity for channels with rapidly decaying correlations such as ISI channels, [20]. Unfortunately, the numerical evaluation of capacity is not an option for channels with long range correlations such as the global jitter channel: due to the strong correlations between *all* signals received at the same time, the calculation of the N-dimensional integral in the right hand side of (37) cannot be reduced to a set of low dimensional problems which makes the numerical evaluation of capacity extremely inefficient.

Fortunately, it turns out that capacity of channel (1) can be calculated asymptotically in the limit of large array size $N >> 1$. The simplification which allows capacity calculation in the large-N limit is easy to understand: on the one hand, the received signals $\mathbf{r}$ are independent conditionally on the value of the amplitude $p = p(J)$. Conditionally on p, channel capacity is given by the well known expression for binary AWGN channels. On the other hand, the value of $p(J)$ can be extracted from the string of $N >> 1$ received signals with relative accuracy of the order of $1/\sqrt{N}$ due to the law of large numbers.

This simple argument suggests the following answer for capacity of the global jitter channel:

$$\lim_{N \to \infty} \frac{1}{N} C_{i.u.d.} = \mathbb{E}_P [C_{AWGN}(p)] \qquad (40)$$

where $C_{AWGN}(p)$ is the capacity of the AWGN channel with fixed signal amplitude p given by:

$$C_{AWGN}(p) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dx \left[\exp\left(-\frac{x^2}{2}\right) \right.$$
$$\left. \times \log_2 \left(1 + \exp\left(fx - \frac{f^2}{2}\right)\right) \right] \qquad (41)$$

where $f = p/\sigma$, and σ is the standard deviation of additive white noise, see [17], Chapter II, for more details. In what follows we will also use the asymptotic expansion of C_{AWGN} valid in the limit of weak noise, $f >> 1$:

$$1 - C_{AWGN}(p) = \frac{\sqrt{2\pi}}{f \ln(2)} \exp\left(-\frac{f^2}{8}\right) \left(1 + O\left(f^{-2}\right)\right). \quad (42)$$

As it turns out, the derivation of (40) is very simple and relies on some basic properties of mutual information which are both fundamental and intuitively obvious. Firstly,

we notice that the mutual information for AWGN channel is maximized for the uniform distribution of inputs. Then a simple rearrangement of terms in (37) gives the following:

$$I(R; A) - N\mathbb{E}_P(C_{AWGN}(p)) = I(P; R) - I(P; (A, R)), \ (43)$$

where $I(P; R)$ is the mutual information between signal amplitude degraded by jitter and the received signal, $I(P; (A, R))$ is the mutual information between signal amplitude and the joint ensemble of channel input and output. Clearly,

$$I(P; (A, R)) \geq I(P; R). \qquad (44)$$

(The information we learn about P from observing A *and* R must be greater or equal to the information about P contained in R.) In case the above argument fails to convince a rigorous-minded reader, here is the proof based on Jensen inequality [10]:

$$I(P; R) - I(P; (A, R)) = \mathbf{E}_{(P,R,A)} \log \frac{d\Pr_{(P,R)}}{d\Pr_{(P,R)|A}} \frac{d\Pr_{R|A}}{d\Pr_R}$$

$$\overset{Jensen}{\leq} \log \mathbf{E}_{(P,R,A)} \frac{d\Pr_{(P,R)}}{d\Pr_{(P,R)|A}} \frac{d\Pr_{R|A}}{d\Pr_R}$$

$$= \log \int_{\Omega_p \times \Omega_r} d\Pr(P, R) = \log 1 = 0.$$

Using (44), relation (43) leads to the following inequality:

$$0 \leq \mathbb{E}_P(C_{AWGN}(p)) - \lim_{N \to \infty} C_{i.u.d} \leq \lim_{N \to \infty} \frac{1}{N} I(P; (A, R)).$$

Therefore, to verify (40) it remains to show that

$$\lim_{N \to \infty} \frac{1}{N} I(P; (A, R)) = 0. \qquad (45)$$

Intuitively, the validity of the above claim is fairly obvious: imagine for example that random variable P is represented by m-bit numbers. Then the mutual information $I(P; (A, R))$ cannot exceed m bits and the limit in the right hand side (45) is trivially zero. For the proof of (45) in full generality, the reader is referred to Appendix B.

The problem of computing capacity of the global jitter channel for large values of N is solved in principle: the right hand side of (40) is a finite-dimensional integral which depends on the probability distribution of $p(J)$. In particular, it is well suited for numerical study. In Figure 7 the capacity of the probe storage channel suffering a global Gaussian jitter is shown for various values of jitter strength σ_J. For $\sigma_J = 0$ the channel is equivalent to the binary AWGN channel with signal amplitude equal to one. As seen from the plot, this channel has the highest capacity. For a fixed SNR, the capacity of global jitter decreases as σ_J increases. For example at the SNR point corresponding to $C_{AWGN} = 0.9$ bits, the capacity of global jitter channel with $\sigma_J = 0.3$ is $C_{i.u.d} = 0.62$ bits, which rules out the use of high rate linear error correction codes for this channel. The good news is that capacity seems to approach 1 bit per channel symbol in the limit of high SNR even in the presence of global jitter. According to Shannon's theorem this means that by reducing channel noise one can read and write information reliably using large parallel probe arrays at a small redundancy cost.
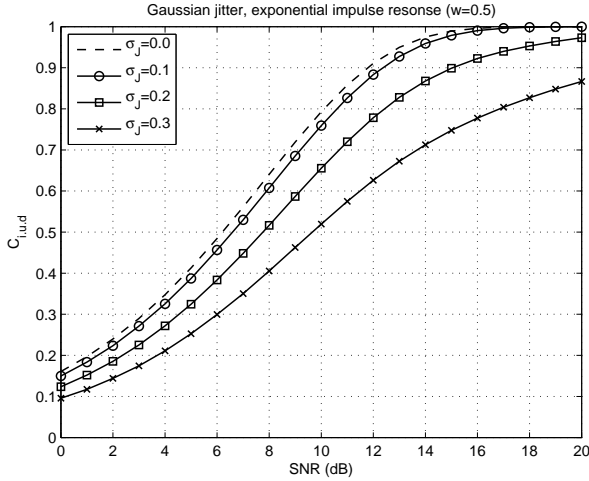
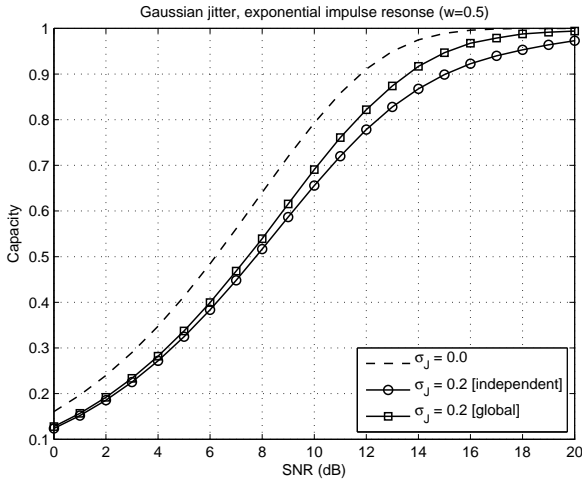Fig. 7. Capacity (i.u.d) for the global jitter channel

Fig. 8. Comparison of capacity (i.u.d) for the global jitter channel against the reference channel

Moreover, the global jitter channel can actually have a larger capacity than a collection of independent probes subject to individual jitter distortions of the same strength: In Figure 8 the capacity of the global jitter channel is compared against the reference channel where each probes suffers an independent Gaussian jitter distortion of the same strength for which the i.u.d capacity can be easily computed numerically. We observe that at low SNR the reference channel has the same capacity than the global jitter channel but at high SNR the global jitter channel actually has a larger capacity! We can explain this phenomenon by the fact that strong correlations between position jitters across the whole array can be used to extract extra information about the hidden parameters of the channel. This extra information can be used for example to build better signal detection algorithms, as discussed in Section III.

To complete our investigation of $C_{i.u.d.}$ we still need to: (i) confirm the numerical observation that capacity approaches 1 bit in the limit of low channel noise; (ii) estimate the corre-

sponding rate of convergence. We will solve these problems for the exponential impulse response assuming weak additive noise and weak jitter, i. e.

$$\sigma << 1, \ \sigma_J << W. \qquad (46)$$

While both of these conditions are reasonable in the context of applications, the weakness of jitter leads to a significant technical simplification of the argument given below. Our starting point is the following bound on the capacity of the AWGN channel:

$$1 - U\frac{\sigma}{p}e^{-\frac{p^2}{8\sigma^2}} \geq C_{AWGN}(p) \geq 1 - Le^{-\frac{p^2}{8\sigma^2}}, \qquad (47)$$

for some positive constants U and L. Eq. (47) results from the straightforward yet tedious analysis of (41). Averaging the above inequality over p we obtain the following bound for the capacity of the global jitter channel:

$$\lim_{N \to \infty} C_{i.u.d.} \geq 1 - L'$$
$$\times \frac{W}{\sigma_J} \left(2^{\frac{W^2}{4\sigma_J^2} - 1} \Gamma \left(\frac{W^2}{2\sigma_J^2} \right) \sigma^{\frac{W^2}{2\sigma_J^2}} + \sqrt{\pi}e^{-\frac{1}{2e^2\sigma^2}} \right) \qquad (48)$$

$$\lim_{N \to \infty} C_{i.u.d.} \leq 1 - U'\frac{W}{\sigma_J}\frac{\sigma^{\frac{W^2}{2\sigma_J^2}}}{\sqrt{\log(1/8\sigma^2)}}I\left(\frac{W}{2\sigma_J} \right), \qquad (49)$$

where L' and U' are positive constants, $\Gamma(x)$ is the Γ-function,

$$I(z) = 8^{\frac{z-1}{2}} \int_0^{1/e} dx x^{z-2} e^{-x^2}.$$

As the right hand side of the lower bound (48) approaches 1 bit in the low noise limit $\sigma = 0$, we conclude that

$$\lim_{\sigma \to 0} \lim_{N \to \infty} C_{i.u.d.} = 1,$$

which confirms the results of numerical simulations. The upper bound (49) shows that the convergence to one is not faster than the power law $\sigma^{\frac{W^2}{2\sigma_J^2}}$, which is much slower than the exponential convergence of $C_{AWGN}(p) \sim 1 - Ae^{-p^2/8\sigma^2}$ to one - only a very 'clean' global jitter channel will have capacity close to the capacity of AWGN channel.

Finally, comparing the lower capacity bound with the upper bound we see that with logarithmic precision the convergence of $C_{i.u.d.}$ for large N is indeed given by the power law:

$$1 - C_{i.u.d.} \sim Const \cdot \sigma^{\frac{W^2}{2\sigma_J^2}}. \qquad (50)$$

Therefore, the capacity approaches 1 bit in exactly the same way as the Reed-Solomon sector error rate vanishes in the limit of zero additive noise, see (36). The power law approach of the channel characteristics to their noiseless limiting values seems to be a feature of the global jitter channel. We will encounter the law (50) again in the next Section, when we compute the Gallager's coding bound for the global jitter channel.

VI. THE RANDOM CODING BOUND FOR THE GLOBAL JITTER CHANNEL.

According to the results of the previous Section, the capacity of the global jitter is positive and even approaches 1 bit in the limit of low noise. Therefore, it follows from Shannon's

theorem that there are error correction codes with rate up to capacity C which can be used to transmit information over the channel with vanishingly small probability of error. Yet as we have seen in Section IV, non-interleaved Reed-Solomon codes exhibit an error floor for any block size and any rate.

Therefore, we are led to the following question: are there *any* error correction block codes (linear or non-linear) which yield a small probability of block error if the code block coincides with the instantaneous output of the probe array?

Let $\mathfrak{C}$ be a code with rate R and block size N (bits) which is also the number of probes in the array. Let $\Pr(SE \mid \mathfrak{C})$ be the probability of block decoding error for this code.

This probability is difficult to compute for any non-trivial code. However, we can use the idea of Gallager [21] and calculate the *average* probability $\overline{\Pr}(SE \mid R, N)$ of block error under the maximum likelihood decoding, where the average is taken over all random codes with a given rate R and block size N. Given $\overline{\Pr}(SE \mid R, N)$ there must exist a code $\mathfrak{C}_0$ with rate R and block size N such that

$$\Pr(SE \mid \mathfrak{C}_0) \leq \overline{\Pr}(SE \mid R, N).$$

Therefore, if we find that $\overline{\Pr}(SE \mid R, N)$ is sufficiently low (e. g. 10^{-12}) for a desired code rate R, we will know that there are codes which can be used to correct errors for information communicated over the global jitter channel.

To perform the calculation of the random coding bound we need to define the space of random codes and a probability measure on this space. Following Gallager, we will construct a random binary code by picking $K = NR$ binary codewords from the N-dimensional binary space $\{0,1\}^N$ independently and uniformly (Gallager ensemble).

Using the sum rule, $\Pr(SE \mid \mathfrak{C})$ can be re-written as follows:

$$\Pr(SE \mid \mathfrak{C}) = \int_0^1 d\mu_P(p) \Pr(SE \mid \mathfrak{C}, p)$$

where $\Pr(SE \mid \mathfrak{C}, p)$ is the probability of block decoding error for the binary AWGN channel with a fixed signal amplitude p. Therefore,

$$\overline{\Pr}(SE \mid R, N) = \int_0^1 d\mu_P(p) \overline{\Pr}(SE \mid p, R, N), \quad (51)$$

where $\overline{\Pr}(Se \mid p, R, N)$ is the random coding bound for the binary AWGN channel with a fixed signal amplitude p. The following set of results can be easily extracted from the original Gallager's paper [21]:

$$\overline{\Pr}(SE \mid p, R, N) \leq e^{-NE(R,p)}, \quad (52)$$

where $E(R, p)$ is the error exponent for the binary AWGN channel. In what follows we do not need an explicit expression for the error exponent, but its basic properties listed below will be important to us:

$$E(R, p) \equiv 0 \text{ for } R \geq C_{AWGN}(p), \quad (53)$$

$$E(R, p) > 0 \text{ for } R < C_{AWGN}(p), \quad (54)$$

$$\frac{\partial E}{\partial R}(R, p) \mid_{R=C_{AWGN}(p)} = 0, \quad (55)$$

$$\frac{\partial^2 E}{\partial R^2}(R, p) \mid_{R=C_{AWGN}(p)} > 0. \quad (56)$$

Using the notion of the error exponent for AWGN channel, we can re-write the random coding bound for the global jitter channel as follows:

$$\overline{\Pr}(SE \mid R, N) \leq \int_0^1 d\mu_P(p) \exp\left[-NE(R,p)\right]. \quad (57)$$

Let p_c be the unique solution to

$$C_{AWGN}(p) = R \quad (58)$$

Then (57) can be rewritten as:

$$\overline{\Pr}(SE \mid R, N) \leq \int_0^{p_c} d\mu_P(p) \exp\left[-NE(R,p)\right]$$
$$+ \int_{p_c}^1 d\mu_P(p) \exp\left[-NE(R,p)\right] \quad (59)$$

The crucial observation is that for $p \leq p_c$, $C_{AWGN}(p) \leq R$. As a result, for $p \leq p_c$, $E(R, p) \equiv 0$ due to (53). Therefore the average probability of decoding error for the global jitter channel is bounded:

$$\overline{\Pr}(SE \mid R, N) \leq \int_0^{p_c} d\mu_P(p)$$
$$+ \int_{p_c}^1 d\mu_P(p) \exp\left[-NE(R,p)\right] \quad (60)$$

The second integral in the right hand side of (60) can be evaluated for $N >> 1$ using Laplace method: due to (54, 55) the main contribution to the integral comes from a small neighbourhood of p_c. It follows from (55,56) that as a function of p, $E(R, p)$ has a non-degenerate critical point at $p = p_c$. A calculation based on the above points yields:

$$\overline{\Pr}(SE \mid R, N) \leq \Pr(p < p_c)$$
$$+ \rho_P(p_c)\sqrt{\frac{\pi}{N\partial_R^2 E(R, p_c)\partial_p C(p_c)^2}} + O(N^{-1}).$$

It is remarkable that in the limit $N \to \infty$ the probability of block decoding error is bounded by a function independent of block size:

$$\lim_{N \to \infty} \overline{\Pr}(SE \mid N, R) \leq \Pr(p < p_c) \quad (61)$$

Therefore, the random coding bound for the global jitter channel does not vanish in the limit of the large block size!

In Figure 9 the random coding bound (61) is shown for various code rates for a global Gaussian positioning error with strength $\sigma_J = 0.2$. The random coding bound for a channel where each probe in an array of $N = 1000$ suffers an independent Gaussian positioning error of the same strength is also given for the same code rates. It is observed that the probability of decoding error (SER) is vastly worse for the case of global positioning errors. For the independent channel the SER decay with a *waterfall* (super-exponential) shape whereas for the global jitter channel the SER decays exponentially with SNR, i. e. exhibits an error floor.

The presence of the error floor in the random coding bound can be demonstrated analytically for Gaussian impulse response and Gaussian jitter in the limit of high SNR and high
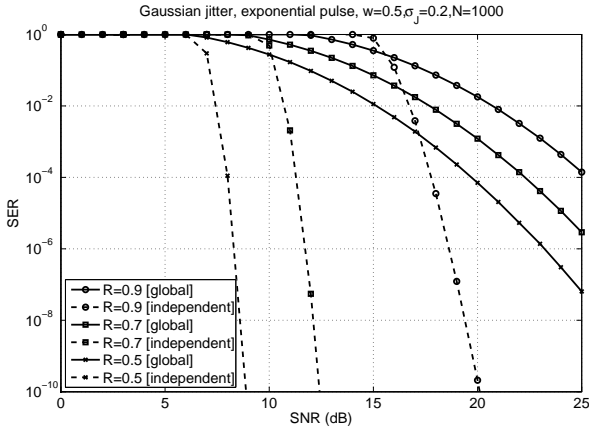
Fig. 9. Random coding bound for the probe storage channel with independent and global jitter for various different rates
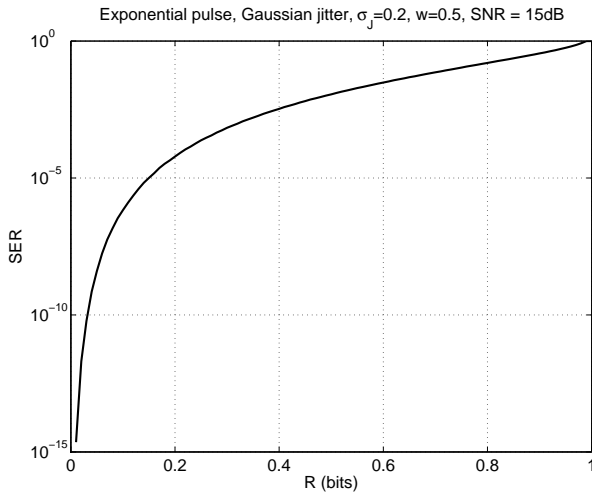

Fig. 10. Behaviour of the random coding bound as a function of rate for the probe storage channel suffering a global Gaussian positioning error

rate codes. The cumulative distribution function $\Pr(p \leq p_c)$ is given by:

$$
\begin{aligned}
\Pr(p \leq p_c) &= \frac{2}{\sqrt{2\pi\sigma_J^2}} \int_{J_c}^{\infty} dJ \exp\left[-\frac{J^2}{2\sigma_J^2}\right] \\
&\leq \exp\left[-\frac{J_c^2}{2\sigma_J^2}\right]
\end{aligned}
\tag{62}
$$

Where $J_c = W\sqrt{\ln(1/p_c)}$ is the positive value of jitter that results in signal loss p. Recall from (2) that W is parameter related to pulse width. Thus the average probability of decoding error is bounded by:

$$
\lim_{N\to\infty} \overline{\Pr}(SE \mid R, N) \leq p_c^{\left(\frac{W^2}{2\sigma_J^2}\right)}
\tag{63}
$$

Furthermore in the limit of weak noise $\sigma \to 0$ and high rate $R \to 1$, it is possible to use the asymptotic expansion for AWGN capacity $C_{AWGN}(p)$ given by (42) to derive an expression for p_c to the logarithmic precision:

$$
p_c \approx \sigma\sqrt{8\ln\left(1/(1-R)\right)}
\tag{64}
$$

Thus for high-rate codes the probability of decoding error can be approximately bounded:

$$
\lim_{N\to\infty} \overline{\Pr}(SE \mid R, N) \lesssim C(\sigma_J, w, R)\, \sigma^{\left(\frac{W^2}{2\sigma_J^2}\right)}
\tag{65}
$$

where $C(\sigma_J, w, R)$ is the σ-independent constant:

$$
C(\sigma_J, w, R) = (-8\ln(1-R))^{\left(\frac{W^2}{4\sigma_J^2}\right)}
\tag{66}
$$

We conclude the average probability of decoding error in the limit of weak noise and high code rate is asymptotically independent of the code block size, exhibits an error floor with an exponent independent of the code rate R and an amplitude with depends on the rate via $log(1-R)$. An identical behaviour has been observed for Reed-Solomon codes in Section IV. Similar conclusions concerning the random coding bound can be reached for low code rates as well, see Fig. 10 where the bound (61) is shown as a function of rate R for a fixed SNR. Note that the average probability of block error approaches zero only in the limit of zero code rate.

The fact that the random coding bound exhibits an error floor behavior identical to that of RS code suggests to us that *all* non-interleaved error correction codes suffer identical performance degradation due to global jitter. This suggestion is confirmed in the following Section.

VII. THE NON-EXISTENCE OF NON-INTERLEAVED ERROR CORRECTION CODES FOR THE GLOBAL JITTER CHANNEL.

The existence of an N-independent error floor in the average probability of block error has the following simple explanation: provided the strength of additive noise is positive and $R > 0$, there is an N-independent critical value of jitter $p_c(R)$ beyond which the conditional channel capacity $C_{AWGN}(p)$ is smaller than the code rate R. Therefore, according to the negative part of Shannon's theorem, information transmission without errors is impossible with N-independent probability $\Pr(p < p_c(R))$.

The above consideration can be turned into a rigorous argument which shows that non-interleaved encoding of the large array's outputs cannot ensure error free information retrieval *no matter what error correction code with positive rate is used*:

Let $\mathfrak{C}(R, N)$ be a block code with rate R and block size N. Clearly,

$$
\Pr(SE \mid \mathfrak{C}(R, N)) \geq \int_0^{p_c} d\mu(p)\,\Pr(SE \mid p, \mathfrak{C}(R, N)).
\tag{67}
$$

Recall that p_c is the unique solution to $R = C_{AWGN}(p)$. Notice that for every p in the region of integration, $R > C_{AWGN}(p)$. By Fano's inequality [22],

$$
\Pr(SE \mid p, \mathfrak{C}(R, N)) \geq \frac{R - C_{AWGN}(p)}{R} - \frac{1}{RN}.
\tag{68}
$$

Therefore,

$$
\begin{aligned}
\Pr(SE \mid \mathfrak{C}(R, N)) &\geq \int_0^{p_c} d\mu(p)\left(1 - \frac{C_{AWGN}(p)}{R}\right) \\
&\quad - \frac{1}{RN}\Pr(p \leq p_c).
\end{aligned}
\tag{69}
$$

Therefore, the probability of block error is bounded below by a constant asymptotically independent of the block size.

Integrating (69) by parts it is easy to show that

$$\lim_{N \to \infty} \Pr(SE \mid \mathfrak{C}(R,N)) \geq \int_0^R \frac{dc}{R} \Pr(C_{AWGN}(p) \leq c), \quad (70)$$

where

$$\Pr(C_{AWGN}(p) < c) = \int_0^1 d\mu(p)\chi(C_{AWGN}(p) < c).$$

Expression (70) proves that the probability of block error is bounded away from zero by an N-independent constant for *any* non-interleaved error correction code with a positive rate R.

Therefore, error-free information storage using large arrays of probes is impossible without interleaving error correction codes between multiple array outputs.

It is worth noting that there is a counterpart of Fano's inequality for the global jitter channel - a non-trivial fact given strong correlations between all probe channels within the array. Using the fact that $\lim_{N \to \infty} C_{i.u.d.} = \mathbb{E}_p(C_{AWGN}(p))$ which we established in Section V we can derive the following weaker bound from the bound (70):

$$\lim_{N \to \infty} \Pr(SE \mid \mathfrak{C}(R,N)) \geq 1 - \frac{\lim_{N \to \infty} C_{i.u.d.}}{R}, \quad (71)$$

which shows the impossibility of error-free information transmission over the global jitter channel using non-interleaved codes with $R > \lim_{N \to \infty} C_{i.u.d.}$.

Finally, let us show that any non-interleaved code used to encode information transmitted over Gaussian global jitter exhibits an error floor identical to Reed-Solomon error floor (32, 36) or the error floor in the random coding bound (63, 65). Note that the capacity of AWGN channel $C_{AWGN}(p)$ is a function of p/σ only,

$$C(p) = F\left(\frac{p}{\sigma}\right),$$

see (41). Substituting expression (3) for the probability measure $d\mu_P(p)$ for Gaussian jitter into the integral in the right hand side of (69) and changing the integration variable $p = \sigma x$ we arrive at the following result:

$$lim_{N \to \infty} \Pr(SE \mid \mathfrak{C}(R,N))$$
$$\geq \frac{\sigma^\gamma}{\sqrt{2\pi\sigma_J^2}} \int_0^{x_c} dx \frac{x^{\gamma-1}}{\sqrt{log\frac{x_c}{x} + log\frac{1}{\sigma x_c}}} \left(1 - \frac{F(x)}{R}\right) \quad (72)$$

where $\gamma = \frac{W^2}{2\sigma_J^2}$ and x_c is the unique positive solution to the equation

$$F(x_c) = R. \quad (73)$$

Notice that $x_c\sigma = p_c < 1$. The following inequality is valid provided the additive noise is weak: If

$$log\frac{1}{x_c\sigma} > 1/2 :$$

and for any $x : \ 0 < x \leq x_c$,

$$\frac{1}{\sqrt{log\frac{x_c}{x} + log\frac{1}{\sigma x_c}}} \geq \frac{1}{\sqrt{log\frac{1}{x_c\sigma}}} \frac{x}{x_c} \quad (74)$$

Using this inequality in (72) we conclude the following:

$$lim_{N \to \infty} \Pr(SE \mid \mathfrak{C}(R,N)) \geq I(R) \frac{\sigma^{\frac{W^2}{2\sigma_J^2}}}{\sqrt{2\pi\sigma_J^2 \log\frac{1}{x_c\sigma}}}, \quad (75)$$

where

$$I(R) = \int_0^{x_c} \frac{dx}{x_c} x^\gamma \left(1 - \frac{F(x)}{R}\right)$$

is σ-independent function of R and $\gamma = \frac{W^2}{2\sigma_J^2}$.

We conclude that *any* non-interleaved error correction code with rate R and a large block size used to transmit information over the global jitter channel exhibits an exponential error floor: $log \Pr(SE)$ is a linear function of signal-to-noise ratio with exponent

$$\gamma = \frac{W^2}{2\sigma_J^2}. \quad (76)$$

The position of the error floor is determined by the function $I(R)$, which is a slow function of code rate, but we do not study it here. Note that the upper bound on the error floor we derived using elementary tools only (Fano's inequality) coincides with the error floor observed directly for Reed-Solomon codes and the random coding bound for the global jitter channel.

VIII. ON GOOD CODES FOR THE GLOBAL JITTER CHANNEL.

In the two previous sections we established the impossibility of error-free information retrieval for large parallel arrays of probes subject to global jitter using non-interleaved error correction codes of *any* positive code rate.

We also established in Section V that the capacity $C_{i.u.d.}$ of global jitter channel is positive and approaches 1 bit as the additive noise strength σ goes to 0. Therefore, by Shannon's theorem there must exist families of error correction codes with rates $0 < R < C_{i.u.d.}$ which ensure vanishingly small probability of block error in the limit of large block sizes.

Can we say anything about the structure of these codes? We understand that non-interleaved codes cannot perform well on the global jitter channel due to rare strong jitter fluctuations leading to small effective AWGN amplitude $p \lesssim \sigma$. Due to this fluctuations N bits of information get lost regardless of error correction code used.

To avoid this we must spread information over many time slices by encoding *blocks* of array outputs. Therefore good error correction codes for the global jitter channel must be interleaved.

Let us estimate the depth of interleaving and the block size of the corresponding codes. Let $\mathbf{r}_{t \in \mathbb{Z}}$ be the time series of array outputs. Recall that each output is an N-dimensional vector, where N is the number of probes in the array. For a feedback-loop based positioning system used in Millipede, jitter random variables J_t are correlated in time. Let L be the correlation length measured in the number of sampling periods. Our previous discussion on the influence of correlations on the performance of error correction codes can be summed up as follows: the probability of decoding error of the block code

of rate $R < C_{i.u.d.}$ and block size B approaches zero as the number of independent groups of samples $\frac{B}{NL}$ goes to infinity. In other words, if NL strongly correlated samples are treated as a single symbol, we get the usual statement of Shannon's theorem for memoryless channels: the probability of block error goes to zero as the number of symbols per block goes to infinity.

The minimal block size of high rate error correction codes used in modern storage devices which achieve the probability of error of the order of 10^{-10} is of the order of 10^3 bits. Clearly, this is the lower bound on the interleaving depth. Therefore we can estimate the block size of the interleaved error correction code for probe storage as

$$B \sim 10^3 NL \qquad (77)$$

Thus for a probe storage device with $N \sim 10^3$ probes in the array the error correction block size should be of the order

$$B \gtrsim 10^6 \text{ bits .}$$

The codes of similar sizes ($5 \cdot 10^5$ bits) are already used in optical storage (Blu-Ray disks), however we thought it unusual to have discovered the need for large block sizes for a non-removable media storage device.

IX. CONCLUSION

In this paper we introduced and analyzed a simple model of the massively parallel probe storage channel suffering from global positioning errors. This model can be viewed as an example of communication channel consisting of $N >> 1$ parallel sub-channels all of which are strongly correlated via a common distortion event.

We solved the problem of optimal signal detection for the global jitter channel. Namely we found a detection algorithm of $O(NlogN)$ complexity such that the estimated sequence converges bitwise to the maximum a posteriori estimate in the limit $N \to \infty$. This is not an entirely trivial result, as in general, the complexity of optimal detection grows exponentially with memory length (for example, the optimal MAP detector for the channel with inter-symbol interference of length I and AWG noise has 2^I states). For the channel at hand the detection algorithm simplifies due to statistical independence of the channel outputs conditional on the value of jitter. This value can be found by analyzing all channel outputs using the law of large numbers.

We analyzed the performance of Reed-Solomon codes applied to the global jitter channel without interleaving channel inputs corresponding to different moments of time. We discovered both numerically and theoretically that in the limit of large probe array, any non-interleaved Reed-Solomon code will exhibit an error floor the position of which is independent of the code's block size N and only weakly dependent on the code rate. This is a surprising result, as the phenomenon of the error floor is usually associated with sub-optimality of the decoding algorithm (e. g. belief propagation) rather than specifics of the channel. For the case of global jitter the origin of the error floor can be traced back to rare instances of strong jitter J such that the signal amplitude $p(J)$ becomes of the same order as channel noise σ.

Motivated by these findings we addressed the following question: are there any good error correction codes for probe storage channels? The answer turned out to be two-fold: firstly, we calculated the capacity $C_{i.u.d.}$ of the probe storage channel and discovered that it does approach one bit per channel symbol in the limit of weak channel noise σ, albeit very slowly: $1 - C_{i.u.d.}$ approaches zero as σ^γ, where the exponent γ is equal to the error floor exponent for Reed-Solomon codes. However, we also found that Gallager's random coding bound for channel encoding which does not interleave between different time slices exhibits exactly the same error floor as Reed-Solomon codes! Moreover, an application of Fano inequality allowed us to prove that any non-interleaved error correction code applied to a massively parallel probe storage channel exhibits an error floor behaviour which is at least as bad as the Reed-Solomon error floor.

With the benefit of hindsight, the appearance of the universal error floor in Reed-Solomon block error rate, Shannon capacity, Gallager's random coding bound and Fano's low bound on sector error rate for an arbitrary fixed code can be related to the statistics of large jitter fluctuations: if $p(J) << \sigma$ then regardless of the code used regions of confuse-ability (a hypersphere of radius σ) for any two codewords intersect and all information and parity is lost!

In this sense the global jitter channel for weak channel noise can be viewed as an effective block-wise erasure channel: either N bits are detected with no errors, or all N bits are lost. For such a channel, it is clearly impossible to improve performance by increasing the number of probes in the array if the encoding block size is kept equal to N. In order to reach Shannon's limit for the global jitter channel it is necessary to spread information between many outputs of the probe array thus mitigating the effects of occasional strong jitter. We estimate the necessary block size of good error correction codes for probe storage channel to be of the order of 10^6 - an unusually large number for a non-removable storage device.

The mathematical results reported in the paper have severe practical implication for probe storage. Either the positioning system must be accurate enough so that global jitter is made very weak or we must interleave the error correction code in the time direction. The former option poses a huge engineering challenge: for indentation sizes of the order of 10 nm, our results suggest that the precision of the positioning system must be significantly better than 1 nm for the effects of global jitter to become insignificant. The latter option also presents significant implementation difficulties especially since jitter is known to be strongly correlated in the time direction and as a result we must interleave very deeply to overcome its effects.

Due to the universal nature of information theoretic performance limits we discovered, we see no easy 'engineering' way around the problem of global jitter. For example one can attempt to re-read the information recovered by the array of probes following a strong jitter event. This is similar to the way off-track errors are dealt with in magnetic hard drives. Our results mean however that the throughput of such a system will be severely degraded due to a large number of necessary re-reads.

From a more theoretical point of view, we developed a

general approach for solving the problem of performance evaluation of channels with long correlations: If the correlations are due to a small number of hidden 'correlation' parameters (e. g. global jitter), we can calculate the conditional capacity, the random coding bound, etc. using well known expressions for memoryless channels and then average over the hidden parameters to obtain the unconditional quantities. The universality of our answers can be explained by the fact that for a large number of sub-channels $N >> 1$ only the values of hidden parameters dictated by the law of large numbers are needed to remove the conditioning.

Given the general nature of our analysis, we hope that results reported in the present paper can be applied to other communications channels with long range correlations between channel outputs. Relevant examples include high density magnetic storage for which long (of the order of several sectors) correlations are present due to off-track errors; optical storage with removable media for which correlations are generated by scratches on media surface; ultra dense NAND flash memory (corresponding to 20 nm and smaller transistor libraries) where long correlations along both bit- and word-lines occur due to read and programme disturbs. An example coming from digital communications is a multiple-input-multiple-output (MIMO) system of receivers and transmitters which is known to be strongly affected by the spatial correlations in the noise vector, [23].

APPENDIX A
THE DERIVATION OF BOUND (12)

Let $\hat{\hat{a}}_t^{(k)}$ and $\hat{a}_t^{(k)}$ be the k-th output of the genie detector (given by equation (5)) and the k-th output of the LLN detector (given by equation (10)) respectively. Then using the triangle inequality and the laws of conditional probability it is possible to show:

$$\left| \Pr\left(\hat{\hat{a}}_t^{(k)} \neq a_t^{(k)} \right) - \Pr\left(\hat{a}_t^{(k)} \neq a_t^{(k)} \right) \right|$$
$$\leq \int d\mu_p(\pi) \Pr\left(\hat{\hat{a}}_t^{(k)} \neq \hat{a}_t^k | p = \pi \right) \qquad (78)$$

The event corresponding to the output of the two detectors being different can be expressed as follows:

$$\Pr\left(\hat{\hat{a}}_t^{(k)} \neq \hat{a}_t^{(k)} | p = \pi \right) =$$
$$\Pr\left(r_t^{(k)} \in \left(\tfrac{p}{2}, T_N \right) \middle| \left| \tfrac{p}{2} - T_N \right| > \epsilon, p = \pi \right)$$
$$\times \Pr\left(\left| \tfrac{p}{2} - T_N \right| > \epsilon \middle| p = \pi \right)$$
$$+ \Pr\left(r_t^{(k)} \in \left(\tfrac{p}{2}, T_N \right) \middle| \left| \tfrac{p}{2} - T_N \right| < \epsilon, p = \pi \right)$$
$$\times \Pr\left(\left| \tfrac{p}{2} - T_N \right| < \epsilon \middle| p = \pi \right) \qquad (79)$$

We proceed to bound each term as follows upper-bound

$$\Pr\left(r_t^{(k)} \in \left(\tfrac{p}{2}, T_N \right) \middle| \left| \tfrac{p}{2} - T_N \right| > \epsilon, p = \pi \right) \leq 1 \quad (80)$$

$$\Pr\left(\left| \tfrac{p}{2} - T_N \right| < \epsilon \middle| p = \pi \right) \leq 1 \quad (81)$$

$$\Pr\left(\left| \tfrac{p}{2} - T_N \right| > \epsilon \middle| p = \pi \right) \leq \frac{\mathrm{Var}\left(\tfrac{p}{2} - T_N | p = \pi \right)}{\epsilon^2} \quad (82)$$

$$\Pr\left(r_t^{(k)} \in \left(\tfrac{p}{2}, T_N \right) \middle| \left| \tfrac{p}{2} - T_N \right| < \epsilon, p = \pi \right)$$
$$\leq \frac{2\epsilon \, e^{-\frac{(\pi/2-\epsilon)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \quad (83)$$

Where (80) and (81) are trivial, (82) is due to the central limit theorem (and convergence of T_N to $p/2$) and (83) is a simple upper bound of the corresponding Gaussian integral. Substituting the resulting upper bound into (78) and integrating over π we find:

$$\left| \Pr\left(\hat{\hat{a}}_t^{(k)} \neq a_t^{(k)} \right) - \Pr\left(\hat{a}_t^{(k)} \neq a_t^{(k)} \right) \right|$$
$$\leq \frac{\mathbb{E}(p^2)/4 + \sigma^2}{N\epsilon^2} + \frac{2\epsilon}{\sqrt{2\pi\sigma^2}} \quad (84)$$

We now minimise with respect to ϵ to obtain the tightest bound possible. The critical value of ϵ is given by:

$$\epsilon_{min} = \left(\frac{(\mathbb{E}(p^2)/4 + \sigma^2)\sqrt{2\pi\sigma^2}}{N} \right)^{1/3} \quad (85)$$

Substituting ϵ_{min} into (84) we arrive at the bound (12).

APPENDIX B
THE CALCULATION OF $\lim_{N\to\infty} \frac{1}{N} I(P;(A,R))$.

We will prove that $\lim_{N\to\infty} \frac{1}{N} I(P;(A,R)) = 0$ not just for Gaussian jitter, but any distribution of $p \in (0,1)$ with finite differential entropy:

$$h = \mathbb{E}_P\left(log\frac{1}{p} \right) < \infty. \quad (86)$$

By definition, the mutual information between signal strength P and the channel's input and output (A,R) is:

$$I(P;\mathbf{R}) = \mathbb{E}_{P,A,R} \log \frac{d\Pr(A,R \mid P)}{d\Pr(A,R)}, \quad (87)$$

Using the independence of signal strength P and channel input A the above expression can be re-written as follows:

$$I(P;\mathbf{R}) = \mathbb{E}_{P,A,R} \log \frac{d\Pr(R \mid A,P)}{d\Pr(R \mid A)}. \quad (88)$$

For Gaussian additive channel noise, the fraction under the sign of the logarithm takes the form

$$F(P,A,R) \equiv \frac{d\Pr(R \mid A,P)}{d\Pr(R \mid A)}$$
$$= \int_0^1 d\mu_P(q) e^{-\frac{N}{2\sigma^2}[2X(P-q)+Y(q^2-p^2)]}, \quad (89)$$

where

$$X = \frac{1}{N} \sum_{k=1}^{N} a_k r_k, \qquad (90)$$

$$Y = \frac{1}{N} \sum_{k=1}^{N} a_k. \qquad (91)$$

Conditionally on P, random variables are X and Y are equal to sums of independent identically distributed random variables and

$$\mathbb{E}(X \mid P = p) = \frac{p}{2}, \qquad (92)$$

$$\mathbb{E}(Y \mid P = p) = \frac{1}{2}. \qquad (93)$$

Therefore, conditionally on P, X and Y converge strongly to their respective expectation values. This motivates the introduction of new random variables strongly converging to zero:

$$X_N = X - \frac{p}{2}, \; Y_N = Y - \frac{1}{2}. \qquad (94)$$

Using X_N, Y_N, the problem of calculating $\lim_{N\to\infty} \frac{1}{N} I(P; (A, R))$ can be formulated as follows: compute

$$L = - \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{P,X_N,Y_N} log F(P, X_N, Y_N), \qquad (95)$$

where

$$F(P, X, Y) = \int_0^1 d\mu_P(q) e^{-\frac{(p-q)^2}{4\sigma^2} - \frac{N}{2\sigma^2}(2X(p-q) + Y(q^2 - p^2))}.$$

Let us fix $\epsilon > 0$. A calculation which essentially repeats the derivation of Chebyshev inequality shows that

$$\Pr(X_N^2 + Y_N^2 \geq \epsilon \mid P = p) \leq \frac{\sigma^2 + 1}{2\epsilon^2 N}. \qquad (96)$$

Now let us re-write L using the partition of unity as the sum of two terms:

$$L = \lim_{N\to\infty} (A_N + B_N), \qquad (97)$$

where

$$A_N = -\frac{1}{N} \mathbb{E}_{P,X_N,Y_N} \mathbf{1}_{X_N^2 + Y_N^2 \leq \epsilon} log F(P, X_N, Y_N), \qquad (98)$$

$$B_N = -\frac{1}{N} \mathbb{E}_{P,X_N,Y_N} \mathbf{1}_{X_N^2 + Y_N^2 > \epsilon} log F(P, X_N, Y_N), \qquad (99)$$

Under the condition that the differential entropy (the logarithmic moment) of $p(J)$ exists, it is easy to show that

$$\mid A_N \mid \leq \frac{3}{2}\sigma^2 \epsilon. \qquad (100)$$

Similarly,

$$\mid B_N \mid \leq \frac{1}{4\sigma^2} \mathbb{E}_{(X_N,Y_N,P)}(2|X_N| + |Y_N|)\mathbf{1}_{X_N^2 + Y_N^2 > \epsilon}$$

$$\leq \frac{1}{4\sigma^2} \sqrt{\mathbb{E}_{(X_N,Y_N,P)}(2|X_N| + |Y_N|)^2} \sqrt{Pr(X_N^2 + Y_N^2 > \epsilon)},$$

where the last line is obtained using Schwarz inequality. Using (96) and the fact that

$$\mathbb{E}(2|X_N| + |Y_N|)^2 = \frac{1}{N} \left(\frac{1}{2} + \sqrt{p^2 + 2\sigma^2} \right)^2$$

we find that

$$|B_N| \leq \frac{\sqrt{\mathbb{E}_P \left(\frac{1}{2} + \sqrt{p^2 + 2\sigma^2} \right)^2}}{N\epsilon}. \qquad (101)$$

The expectation value in the above expression always exists as the random variable P is bounded. Combining (100) and (101) we find that for any $\epsilon > 0$ and any N,

$$|A_N + B_N| \leq \frac{3}{2}\sigma^2 \epsilon + \frac{C}{\epsilon N}$$

for some ϵ, N- independent constant C. Choosing $\epsilon = N^{-1/2}$ and taking the limit $N \to \infty$, we find that

$$L = \lim_{N\to\infty} (A_N + B_N) = 0.$$

ACKNOWLEDGMENT

REFERENCES

[1] Binnig G, Quate CF, Gerber C., *Atomic force microscope*, Phys Rev Lett., **56**(9), pp. 930–933 (1986).

[2] H. Bhaskaran, A. Sebastian, A. Pauza, L. Wang, D. Wright, M. Despont, H. Pozidis, E. Eleftheriou, *Recent Advances in Conduction Mode Scanning Probes for Phase Change Probe Storage Applications*, Proc. European Symposium on Phase Change and Ovonic Science "EPCOS 2009", Aachen, Germany, pp. 49–56, (September 2009).

[3] H.J. Mamin, R.P. Ried, B.D. Terris, D. Rugar, B.W. Chui, and T.K. Kenny, *High density data storage based on the atomic force microscope*. In Non- volatile Memory Technology Conference, 1998. Proceedings. Seventh Biennial IEEE, p. 65, (1998).

[4] Campardo, Giovanni; Tiziani, Federico; Iaculo, Massimo (Eds.), *Memory Mass Storage*, Springer (2011).

[5] E. Eleftheriou, T. Antonakopoulos, G. K. Binnig, G. Cherubini, M. Despont, A. Dholakia, U. Drig, M. A. Lantz, H. Pozidis, H. E. Rothuizen, and P. Vettiger, *Millipede a MEMS-based Scanning-Probe Data-Storage System*, IEEE Transactions on Magnetics, vol. **39**, no.2, pp. 938–945 (March 2003).

[6] H. Pozidis, G. Cherubini, A. Pantazi, A. Sebastian and E. Eleftheriou, *Channel Modelling and Signal Processing for Probe Storage Channels*, IEEE J. Sel. Ar. Commun., vol. **28**, no. 2, pp. 143–157 (Feb. 2010).

[7] M. Despont, J. Brugger, U. Drechsler, U. Drig, W. Hberle, M. Lutwyche, H. Rothuizen, R. Stutz, R. Widmer, H. Rohrer, G. Binnig and P. Vettiger, *VLSI-NEMS Chip for Parallel AFM Data Storage*, Technical Digest MEMS '99, 12th IEEE Int. Conf. on Micro Electro Mechanical Systems (IEEE, Piscataway, 1999), pp. 564–569.

[8] P. Vettiger, G. Cross, M. Despont, U. Drechsler, U. Drig, B. Gotsmann, W. Hberle, M.A. Lantz, H.E. Rothuizen, R. Stutz, and G.K. Binnig *The "Millipede" – Nanotechnology Entering Data Storage*, IEEE Trans. Nanotechnology, vol. 1, no. 1, pp. 39-55 (March 2002).

[9] A. Sebastian, A. Pantazi, and H. Pozidis, *Jitter investigation and performance evaluation of a small-scale probe storage device prototype*, IEEE Global Communications Conference, November 2007, pp. 288–293.

[10] R. Durret, *Probability: Theory and Examples*, 3rd edition, Duxbury Press (2004).

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall (1998).

[12] H. Pozidis, et al., *Signal processing for probe storage*, In Proc. IEEE ICASSP 2005, pp. 745-748, Philadelphia, PA, U.S.A. (2005).

[13] T.P. Parnell, H. Pozidis, and O.V. Zaboronski, *Performance evaluation of the probe storage channel*, In Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE, pp. 1-6 (2009).

[14] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, 2nd edition, Prentice Hall: Englewood Cliffs, NJ (2004).

[15] M. Evans, N. Hastings, B. Peacock, *Statistical Distributions*, 3rd edition, New York: Wiley, pp. 134-136 (2000).

[16] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd ed. Springer, New York (1998).

[17] David J. C. MacKay, *Information theory, inference, and learning algorithms*, Cambridge University Press, 1st edition (2003).

[18] A. Erdelyi, *Asymptotic Expansions*, Dover (1956).

[19] C. E. Shannon, *A mathematical theory of communication*, SIGMOBILE Mob. Comput. Commun. Rev., **5**(1), pp.3–55 (2001).

[20] D. Arnold and H.-A. Loeliger, *On the information rate of binary-input channels with memory*, in Proc. ICC, vol. **9**, Helsinki, Finland, pp. 2692-2695 (2001).

[21] R. Gallager, *A simple derivation of the coding theorem and some applications*, Information Theory, IEEE Transactions on, **11**(1), pp. 3-18 (1965).

[22] R. Fano, *Transmission of information; a statistical theory of communications*, Cambridge, Mass., M.I.T. Press (1961).

[23] J. Salz, *Digital transmission over cross-coupled linear channels*, At&T Technical Journal, vol. **64**, no. 6, pp. 11470–1159 (July-August 1985).